



①9 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENTAMT

⑫ **Offenlegungsschrift**
⑩ **DE 44 12 309 A 1**

⑤1 Int. Cl. 6:
G 10 L 3/00

②1 Aktenzeichen: P 44 12 309.4
②2 Anmeldetag: 9. 4. 94
④3 Offenlegungstag: 12. 10. 95

DE 44 12 309 A 1

⑦1 Anmelder:
Telenorma GmbH, 60326 Frankfurt, DE

⑦2 Erfinder:
Euler, Stephan, Dr., 63512 Hainburg, DE; Zinke,
Joachim, Dr., 61194 Niddatal, DE

⑤4 Verfahren zur Verbesserung der Spracherkennung unter Berücksichtigung verschiedener digitaler Codierbedingungen

⑤7 Um eine optimale Spracherkennung zu ermöglichen, wird zunächst dasjenige Codierverfahren ermittelt, welches bei einer über ein Kommunikationsnetz ankommenden Sprachäußerung angewendet worden ist. Dieses Ermittlungsergebnis soll dazu dienen, die Qualität einer Spracherkennung wesentlich zu erhöhen.

Es sind mehrere Spracherkenner vorgesehen, wovon jeder einzelne für ein ganz bestimmtes Codierverfahren zuständig ist. Die Wortmodelle eines jeden Spracherkenners wurden aus Sprachäußerungen erzeugt, die nach dem betreffenden Verfahren codiert waren. Die aus einer aktuellen, zu erkennenden Sprachäußerung errechneten Merkmalsvektoren werden einem Codeerkenner angeboten, in welchem für jedes Codierverfahren ein aus den Merkmalsvektoren der entsprechenden Referenzmuster errechneter Mittelwert abgespeichert ist. Es findet ein Vergleich dieser Mittelwerte mit den Merkmalsvektoren der zu erkennenden Sprachäußerung statt. Mit der sich dabei ergebenden geringsten Differenz ist dann dasjenige Codierverfahren erkannt worden, mit dem die zu erkennende Sprachäußerung codiert wurde. Daraufhin wird derjenige von mehreren Spracherkennern aktiviert, der für das betreffende Codierverfahren zuständig ist.

Durch schnelle Ermittlung desjenigen Codierverfahrens, welches bei einer ankommenden, zu erkennenden Sprachäußerung angewendet wurde, wird durch den darauf spezialisierten Spracherkenner eine hohe Qualität der Erkennungsergebnisse erreicht, die auch bei Codierverfahren mit niedriger ...

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

BUNDESDRUCKEREI 08. 95 508 041/285

5/27

DE 44 12 309 A 1

Die Erfindung betrifft ein Verfahren zur Verbesserung der Spracherkennung unter Berücksichtigung verschiedener digitaler Codierbedingungen nach dem Oberbegriff des Patentanspruchs 1.

Ein Meßverfahren zum Beurteilen der Güte von Sprachcodierern und/oder Übertragungsstrecken ist beschrieben in der DE 37 08 002 A1. In der Beschreibungseinleitung (Spalte 2) ist angegeben, daß die Art des Codiervorgangs und auch die Beschaffenheit einer Übertragungsstrecke einen großen Einfluß haben auf die Sprachgüte und die Verständlichkeit der an einen Empfänger ankommenden Sprachinformation. In dieser Schrift wird ein Meßverfahren angegeben, wobei mit Hilfe eines Spracherkenners oder eines Sprecherkenners festgestellt wird, wie groß die Wahrscheinlichkeit ist, daß abgegebene Sprachäußerungen richtig erkannt werden. Hierbei werden verschiedene Sprachcodierer oder Übertragungsstrecken in den Sprechweg eingeschleift, so daß festgestellt werden kann, wie groß der Anteil der richtig erkannten Sprachäußerungen bei unterschiedlichen Codiervorgängen ist. Mit einem derartigen Verfahren kann festgestellt werden, welches der verschiedenen Codiervorgänge für eine elektronische Spracherkennung am geeignetsten ist und bei welchen Codiervorgängen eine mindere Qualität der Spracherkennung zu erwarten ist. Hierfür wird ein einziger Spracherkenner benutzt, der vorgegebene Sprachäußerungen übermittelt bekommt und diese erkennt. Dabei werden jeweils die vorher bekannten wahren Bedeutungen mit den erkannten Ergebnissen verglichen. Eine Anpassung des Spracherkenners an ein Codiervorgang findet nicht statt.

Aus der DE 43 25 404 A1 ist ein Verfahren zum Ermitteln und Klassifizieren von Störgeräuschtypen bekannt. Damit soll erreicht werden, daß ein Spracherkenner optimal an ein erkanntes Störgeräusch angepaßt werden kann, um die Spracherkennung zu optimieren. Zu diesem Zweck sind sogenannte Codebuch-Speicher vorgesehen, welche die Referenzmuster der jeweils zu erkennenden Sprachäußerungen enthalten. Außer einem Standardcodebuch, in welchem die Referenzmuster der reinen Sprache enthalten sind, ist für jeden zu erkennenden Geräuschtyp ein eigenes Codebuch erforderlich. Diese Referenzmuster müssen zusammen mit dem jeweiligen Geräusch durch einen Trainingsvorgang erstellt werden. Eine zu erkennende Sprachäußerung wird einer Sprachanalyseeinheit angeboten, die Merkmalsvektoren bildet und die Häufigkeitsverteilungen, die aus einer Vektorquantisierung resultieren, ebenfalls ermittelt. Es müssen nun umfangreiche Vergleichsrechnungen stattfinden, um die abgelegten Indexdaten mit den aus der Vektorquantisierung der eingegebenen Sprachäußerung sich ergebenden Häufigkeitsverteilungen zu vergleichen. Der sich bei diesem Rechenvorgang ergebende Extremwert der Differenzen kennzeichnet einen gefundenen Störgeräuschtyp. Daraufhin wird dann dem Spracherkenner das zu diesem Störgeräuschtyp passende Codebuch zugeschaltet, womit dann die Spracherkennung stattfinden kann.

Obwohl bei diesem Verfahren bereits gegenüber vorbekannten Anordnungen eine Reduzierung der Speicherkapazität und der Rechenleistung erreicht wurde, ist der Aufwand dennoch relativ hoch. Außerdem werden bei einem derartigen Verfahren Einflüsse von Codiervorgängen, wie sie auf Übertragungsstrecken eingesetzt werden, nicht berücksichtigt.

Die Aufgabe der Erfindung besteht darin, ein Verfahren anzugeben, mit dem dasjenige von mehreren bekannten Codiervorgängen ermittelt werden kann, welches bei einer zu erkennenden aktuellen Sprachäußerung angewendet worden ist. Daraufhin soll ein Spracherkenner eingesetzt werden, dessen zugehörige Referenzmuster mit diesem Codiervorgang erstellt wurden, damit optimale Ergebnisse erreicht werden, wenn über ein Kommunikationsnetz übertragene Sprache erkannt werden soll.

Zur Lösung dieser Aufgabe ist eine Merkmalskombination vorgesehen, wie sie im Patentanspruch 1 angegeben ist.

Damit wird in vorteilhafter Weise erreicht, daß ein Codiervorgang, welches bei einer über ein Kommunikationsnetz ankommenden, zu erkennenden Sprachäußerung angewendet wurde, schnell gefunden wird. Daraufhin wird derjenige Spracherkenner angesteuert, welcher für dieses Codiervorgang optimal ausgelegt ist. Somit ergibt sich eine hohe Qualität bei der Spracherkennung, die auch bei Codiervorgängen mit niedriger Bitrate ausreichend ist.

Weiterbildung der Erfindung ergeben sich aus den Unteransprüchen.

Ein Ausführungsbeispiel der Erfindung wird nachfolgend anhand einer Zeichnung näher erläutert. In der Zeichnung ist schematisch dargestellt, wie eine von einem Mikrofon M aufgenommene sprachliche Äußerung über ein Kommunikationsnetz KN an den Eingang E einer Sprachanalyse-Einrichtung SPA gelangt. Auf dem Übertragungsweg innerhalb des Kommunikationsnetzes KN sind üblicherweise Codiereinrichtungen CD vorgesehen. Es kann nicht unbedingt davon ausgegangen werden, daß die am Eingang E einer Sprachanalyse-Einrichtung SPA anstehende, zu erkennende sprachliche Äußerung immer mit dem gleichen Codiervorgang angeboten wird. Um eine hohe Qualität bei der Spracherkennung zu erreichen, soll zunächst herausgefunden werden, welches Codiervorgang innerhalb des Kommunikationsnetzes KN angewendet wurde.

In der Sprachanalyse-Einrichtung SPA wird eine zu erkennende Sprachäußerung, beispielsweise ein Wort, in zeitliche Abschnitte zerlegt, wovon in bekannter Weise Merkmalsvektoren MV errechnet werden. Diese Merkmalsvektoren MV werden der Reihe nach in einem Merkmalsvektorspeicher MVSP abgelegt. Von dort aus werden sie einem Codeerkenner CE angeboten, welcher diese Merkmalsvektoren MV mit Mittelwerten MW1 bis MWn vergleicht, welche aus Merkmalsvektoren gebildet wurden, die beim Trainieren von Wortmustern entstanden sind. Jeder einzelne dieser Mittelwerte MW1 bis MWn stammt dabei aus Merkmalsvektoren, welche einem bestimmten Codiervorgang zugeordnet sind, das heißt, diese Merkmalsvektoren wurden von Sprachäußerungen gewonnen, die unter Einsatz des betreffenden Verfahrens codiert wurden. Beim Vergleichen der von einer aktuellen Sprachäußerung gewonnenen Merkmalsvektoren MV mit den im Codeerkenner CE abgelegten Mittelwerten MW1 bis MWn ergeben sich mehr oder weniger große Differenzen. Dabei wird herausgefunden, bei welchem Mittelwert MW1 bis MWn sich die geringste Differenz ergibt.

Da derjenige Mittelwert, z. B. MW1, bei dem sich die geringste Differenz beim Vergleichen mit den Merkmalsvektoren MV der aktuellen Sprachäußerung ergibt, einem Codiervorgang zugeordnet ist, ist dieses somit als dasjenige erkannt worden, welches innerhalb des Kommunikationsnetzes KN angewendet wurde. Wenn

der Codeerkenner CE ein derartiges Ergebnis, das heißt die geringste Differenz, festgestellt hat, so wird ein Schmittmittel SM angesteuert, um einen von mehreren Spracherkennern SE1 bis SEN anzusteuern. Unter der Annahme, daß im Codeerkenner CE die geringste Differenz zwischen dem Mittelwert MW1 und den aus einer aktuellen Sprachäußerung ermittelten Merkmalsvektoren MV erkannt wird, wird über das Schmittmittel SM der Ausgang des Merkmalsvektorspeichers MVSP mit dem Eingang des Spracherkenners SE1 verbunden. Die in diesem Spracherkennung SE1 abgelegten Referenzmuster für zu erkennende Worte sind in bekannter Weise während einer Trainingsphase entstanden, wobei das betreffende Codierungsverfahren, beispielsweise ADPCM, angewendet wurde. Dieser Spracherkennung SE1 liefert deshalb bessere Ergebnisse als einer der anderen Spracherkennung SE2 bis SEN liefern würde. Somit erscheint am Ausgang A bei jedem Codiervorgang, für welches jeweils ein Spracherkennung SE1 bis SEN vorgesehen ist, ein optimales Erkennungsergebnis.

Die mit diesem Verfahren erzielbaren Ergebnisse können noch weiter gesteigert werden, wenn innerhalb des Codeerkenners CE nicht nur einzelne Mittelwerte MW1 bis MWn gespeichert sind, sondern auch die Kovarianzen, welche sich aus den beim Training erzeugten Merkmalsvektoren ergeben. Diese Kovarianzen sind dann in einem Speicher des Codeerkenners CE bezogen auf das jeweilige Codiervorgang abgelegt. In diesem Fall werden von den Merkmalsvektoren MV der aktuellen Sprachanalyse die Werte von Gaußschen Verteilungskurven produziert, so daß einzelne Dichtewerte miteinander verglichen werden können. Mit diesem erhöhten Aufwand kann die Erkennungssicherheit wesentlich gesteigert werden.

Außerdem ist es möglich, mehrere aus ein und derselben Verbindung stammende Sprachäußerungen im Codeerkennung CE nach der vorbeschriebenen Weise zu untersuchen, wobei jede sich bei einer einzelnen Sprachäußerung ergebende Differenz zwischen deren Merkmalsvektoren MV und den Mittelwerten MW1 bis MWn zwischengespeichert wird. Diese einzelnen Differenzwerte werden dann bezogen auf jeden einzelnen Mittelwert MW1 bis MWn getrennt aufsummiert, so daß der dann sich ergebende geringste Summenwert das angewandte Codiervorgang genauer bestimmen kann. Somit kann einer der Spracherkennung SE1 bis SEN mit größerer Sicherheit als der am besten passende ausgewählt werden.

daß eine zu erkennende Sprachäußerung zunächst einer Sprachanalyse-Einrichtung (SPA) zugeführt wird, welche daraus Merkmalsvektoren (MV) erzeugt, die anschließend in einen Merkmalsvektorspeicher (MVSP) gelangen,

daß die Merkmalsvektoren (MV) einem Codeerkennung (CE) angeboten werden, in welchem für jedes Codiervorgang ein aus den Merkmalsvektoren der entsprechenden Referenzmuster errechneter Mittelwert (MW1 bis MWn) abgespeichert ist, daß die Merkmalsvektoren (MV) der zu erkennenden Sprachäußerung mit jedem dieser Mittelwerte (MW1 bis MWn) verglichen werden, wobei die geringste Differenz über alle Merkmalsvektoren (MV) einer Sprachäußerung ermittelt wird, welche dann dasjenige Codiervorgang kennzeichnet, mit dem die zu erkennende Sprachäußerung codiert wurde,

und daß dann derjenige von mehreren Spracherkennung (SPE1 bis SPEn) mit dem Merkmalsvektorspeicher (MVSP) verbunden wird, der für das erkannte Codiervorgang zuständig ist, um die eigentliche Spracherkennung durchzuführen.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß der Codeerkennung (CE) beim Vergleichen der Mittelwerte (MW1 bis MWn) mit den Merkmalsvektoren (MV) auch deren Kovarianzen in Form einer Gauß-Verteilung berücksichtigt, indem einzelne Dichtewerte miteinander verglichen werden.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß bei mehreren untersuchten Sprachäußerungen die bei vorangegangenen Ermittlungen erkannten Differenzen berücksichtigt werden, indem die einzelnen sich ergebenden Differenzen zwischen den Merkmalsvektoren (MV) und den Mittelwerten (MW1 bis MWn) für jedes Codiervorgang aufsummiert werden, wobei die geringste Summe das erkannte Codiervorgang kennzeichnet.

4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß elektronische Schmittmittel (SM) vorgesehen sind, womit der Merkmalsvektorspeicher (MVSP) mit dem entsprechenden Spracherkennung (z. B. SPE1) verbindbar ist, und daß diese Schmittmittel (SM) vom Codeerkennung (CE) direkt gesteuert werden.

Hierzu 1 Seite(n) Zeichnungen

Patentansprüche

1. Verfahren zur Verbesserung der Spracherkennung unter Berücksichtigung verschiedener digitaler Codierungsbedingungen mit niedriger Bitrate, wobei durch Training mit mehreren Sprechern gewonnene Wortmodelle abgespeichert sind, die als Referenzmuster zum Vergleich mit zu erkennenden Worten dienen, damit die Vergleichsergebnisse mit einer hohen Erkennungswahrscheinlichkeit vorliegen, **dadurch gekennzeichnet**, daß mehrere Spracherkennung (SPE1 bis SPEn) vorgesehen sind wovon jeder einzelne für ein ganz bestimmtes Codiervorgang, z. B. PCM mit 64 kBit/s, ADPCM, CELP nach CCITT G728, zuständig ist, wobei die darin befindlichen Referenzmuster der Wortmodelle mit dem zugehörigen Codiervorgang erzeugt wurden,

